



**National Aeronautics and  
Space Administration**

**Jet Propulsion Laboratory**  
California Institute of Technology  
Pasadena, California

## **NASA Sea Level Change Portal – it is not just another portal site**

**Thomas Huang**

Data Scientist | Principal Investigator | Technologist | Architect  
[thomas.huang@jpl.nasa.gov](mailto:thomas.huang@jpl.nasa.gov)

Jet Propulsion Laboratory  
California Institute of Technology  
4800 Oak Grove Drive, Pasadena, CA 91109-8099, U.S.A.



## Principal Investigator

NASA AIST OceanWorks – Ocean Science Platform on Cloud

## Project Technologist

NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC)

## Co-Investigator and Architect

NASA Sea Level Change Portal

## Architect

CEOS Ocean Variables Enabling Research and Application for GEOS (COVERAGE)

## Architect

Tactical Data Science Framework for Naval Research

## Cluster Chair

Federation of Earth Science Information Partners (ESIP) Cloud Computing

## Previously Principal Investigator / Co-Investigator

Several NASA-funded Big Data Analytic Projects – Big Data Analytics on the Cloud, Anomaly Detection, In Situ and Satellite Matchup, Search Relevancy, and Quality Screening

## Goal for the NASA Sea Level Change Team

- Determine how much will sea level rise by [2100]?
- What are the key sensitivities?
- Where are the key uncertainties?
- Where are the key Observables? Model Improvements

## Goals for the NASA Sea Level Change Portal

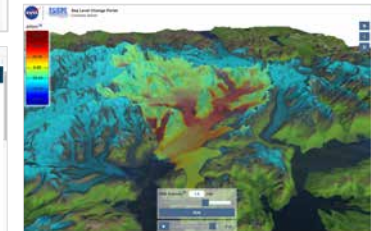
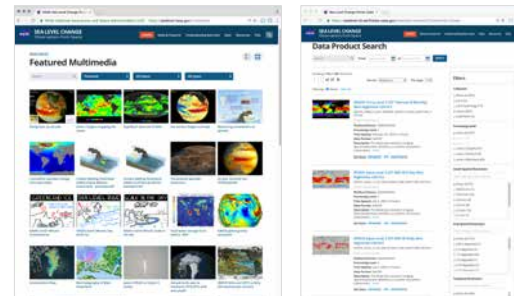
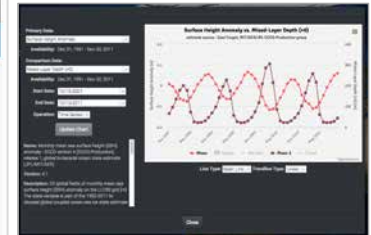
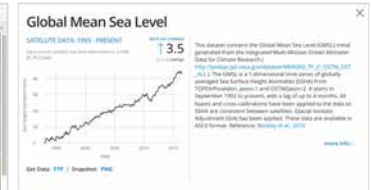
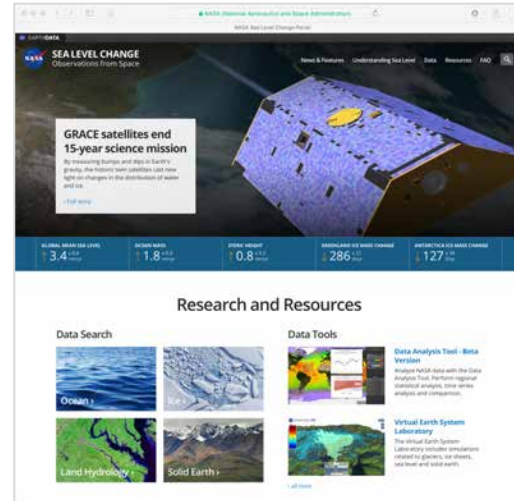
- Provide scientists and the public with a “one-stop” source
- Provide current sea level change information and data
- Provide interactive tools for analyzing and viewing regional data
- Provide virtual dashboard for sea level indicators
- Provide latest news, quarterly report, and publications
- Provide ongoing updates through a suite of editorial products

## Requires

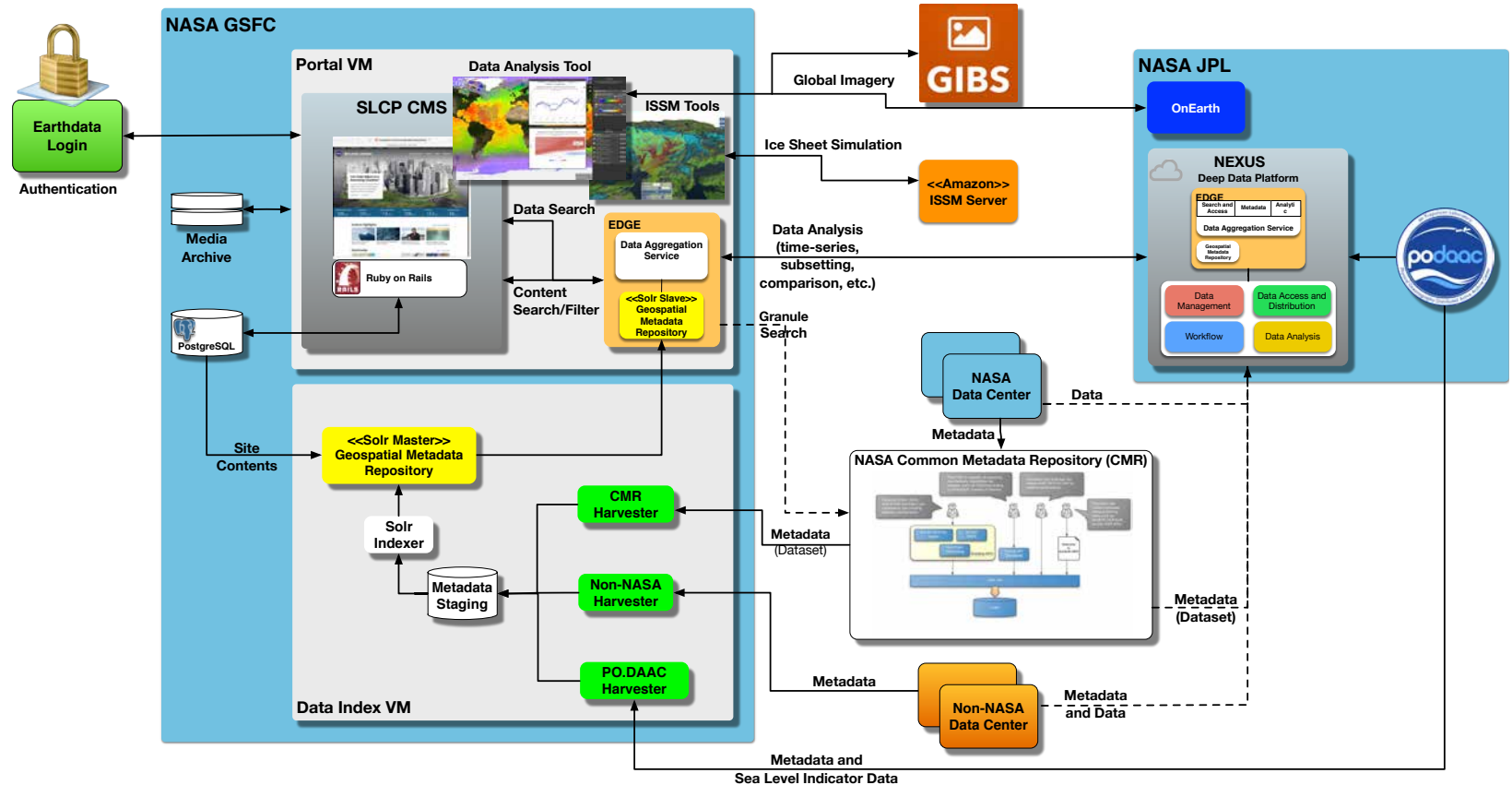
- Interdisciplinary collaboration
- Connect disciplines and evaluate dependencies

## Sea Level Change Portal facilitates

- Easy interdisciplinary data comparison
- Access to latest news and information
- Collaboration (data and information exchange)



# Bringing It Together



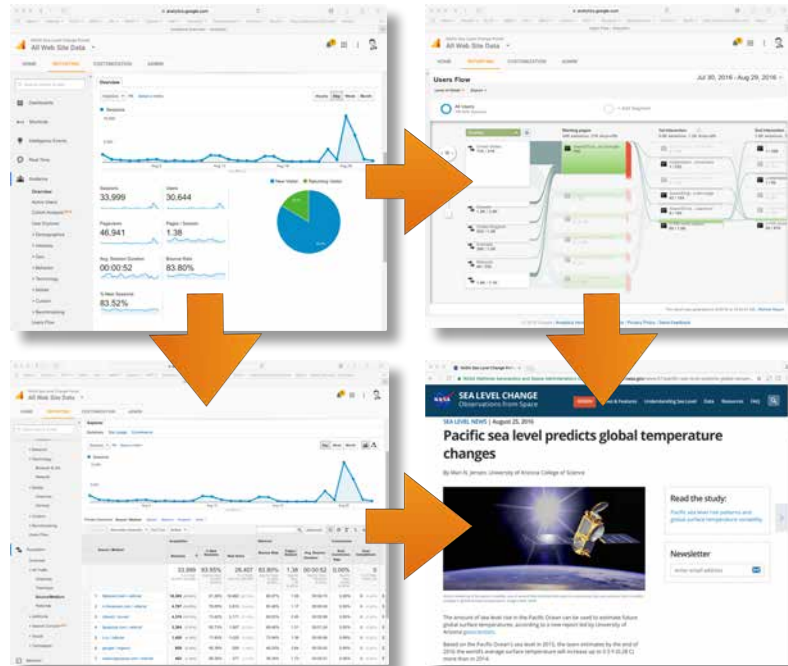
# Analyze User Interactions

Guide website layout

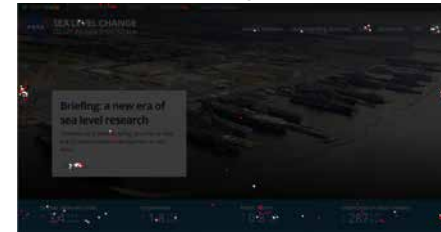
Determine effectiveness of articles and contents

Identify popular media outlines

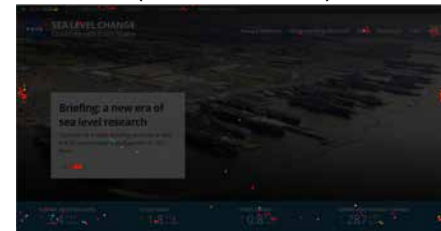
New and returning users



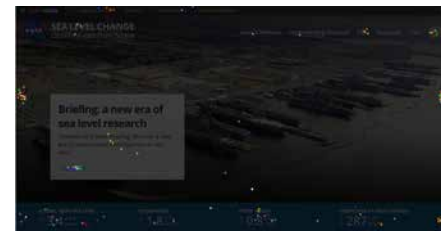
New vs. Returning



Referrer (Media Outlets)



Time to Click



# UX Assessment

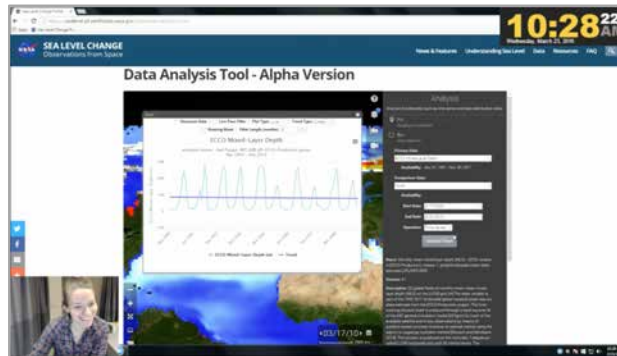
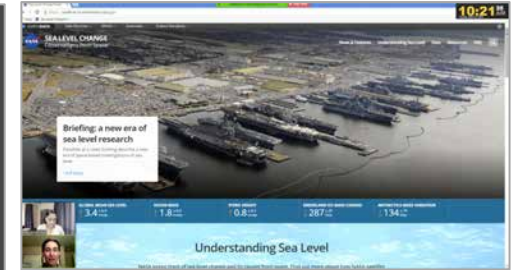
- **Understand** how and for what purposes users obtain sea level change data and information
- **Describe** users' pain points and unmet needs for extracting, visualizing, comparing and analyzing sea level change data
- **Investigate** users' current practices, interests and preferences for engaging in interaction with other sciences and the public about sea level change
- **Identify** opportunities for enhancing cross-disciplinary collaborative activities on the web portal.



## Executive summary

- ① The data analysis tool was a delighter, but there are opportunities to increase its value by adding labels and improving navigability within the tool.
- ② The ice sheet modeling tool was perceived as too rudimentary. There are opportunities for additional interactions and functionality to make it more useful for sea level change scientists.
- ③ Participants were not always successful in finding dataset or publication results that were relevant to their search terms, which raised questions about comprehensiveness. There is an opportunity to increase the amount of information available in the two databases.
- ④ There are opportunities to ease navigation and feature discovery by optimizing menu labels and homepage content on the portal.

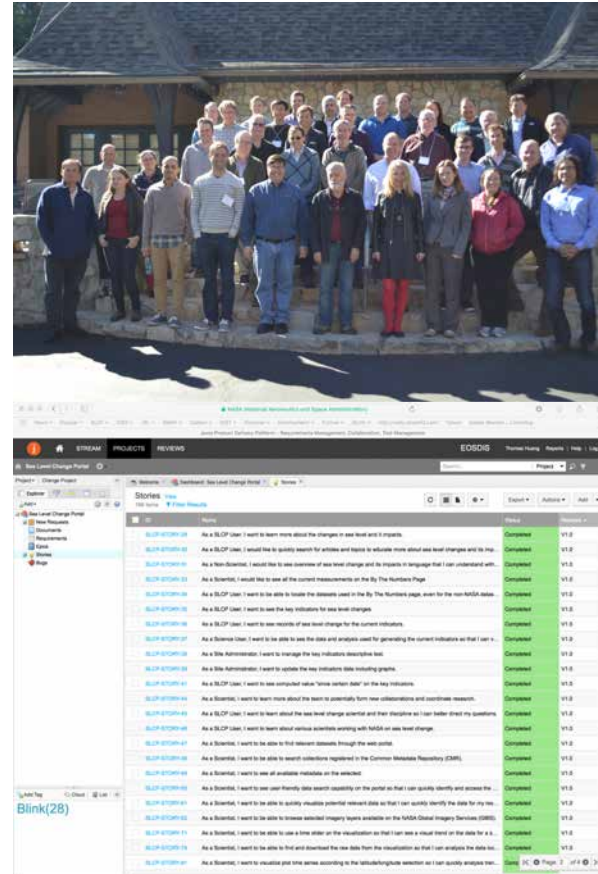
14





# Requirement Management

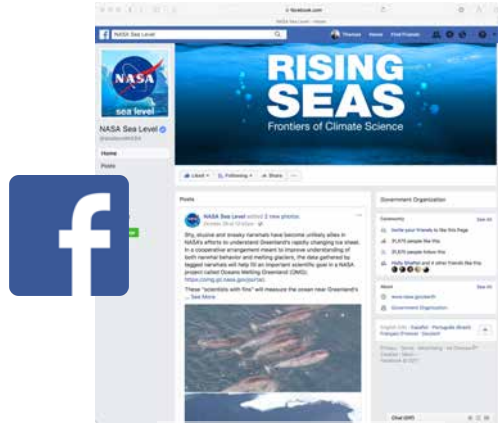
- User Stories: 157 stories
  - UI/UX assessment
  - NASA Sea Level Science Team
  - User feedbacks
- Internal team review before each new development cycle
- Categorize
- Prioritize
- Project Management Approval
- Schedule
- Weekly development team review



2015  
N-SLCT Meeting

JAMA

# Web, Social Media, and Headliners



**Over 31,000  
Followers**



**Over 23,000  
Followers**

**373,000**  
monthly page views

**172,000**  
sessions

**143,000**  
users

## HEADLINES

**“NASA Sea Level Change Website Offers Everything You Need to Know About Climate Change”**  
Tech Times

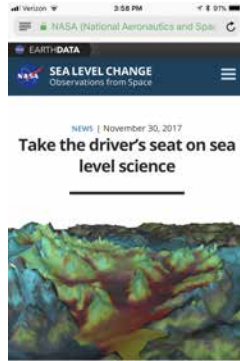
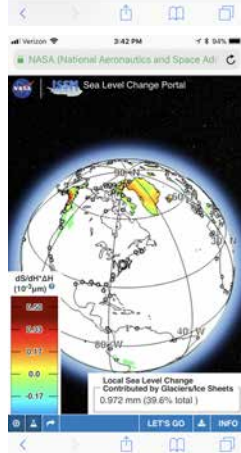
**“NASA’s New Sea Level Site Puts Climate Change Papers, Data, and Tools Online”**  
Tech Crunch



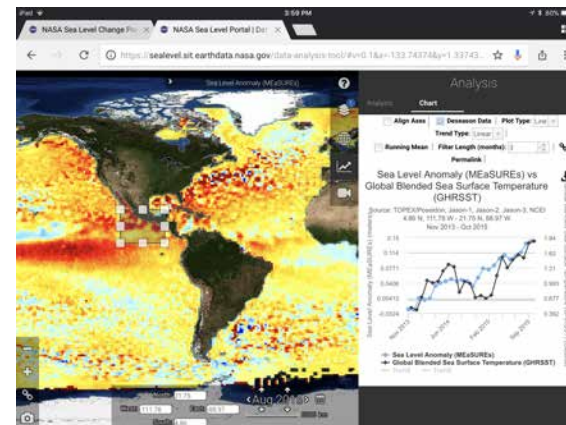
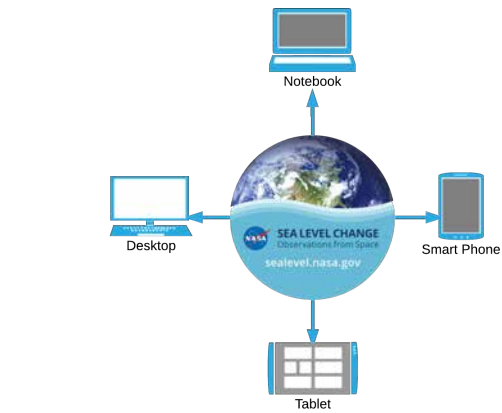
# Mobile Friendly



## Research and Resources

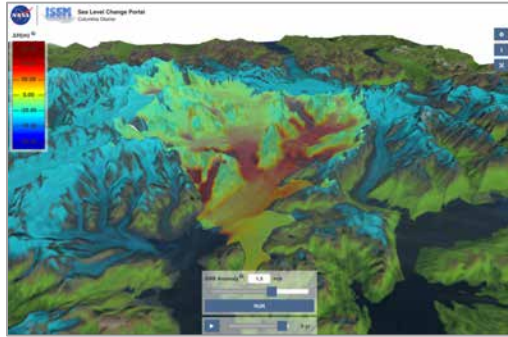


The Gravity Recovery and Climate Experiment (GRACE) twin satellites, launched 17 March 2002, are making detailed measurements of Earth's gravity field and improving investigations about Earth's water.

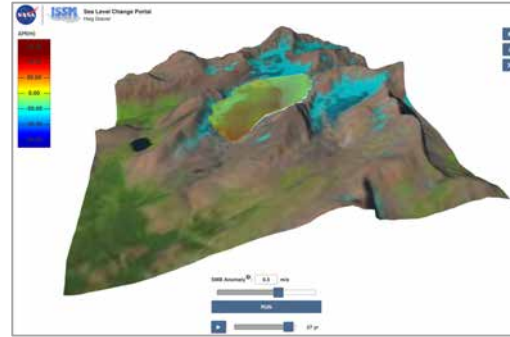


# Virtual Earth System Laboratory (VESL)

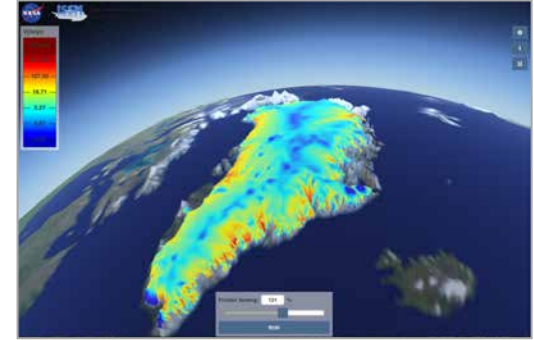
Suite of Interactive Simulation Tools for Glaciers, Ice Sheets, Sea Level, and Solid Earth  
 Simulation Service on Amazon Cloud



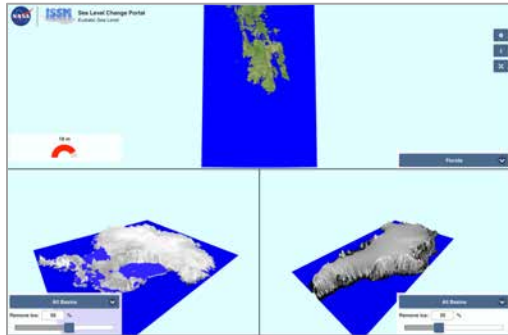
Columbia Glacier



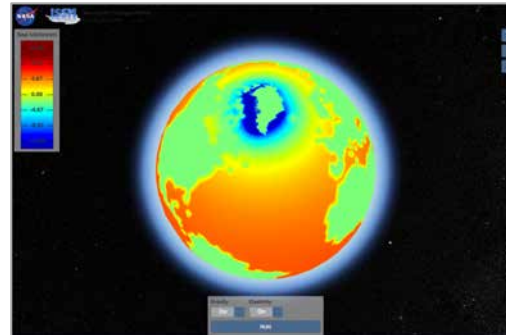
Haig Glacier



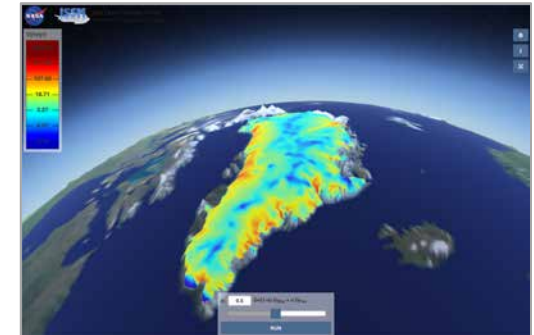
Greenland Basal Friction



Eustatic Sea Level

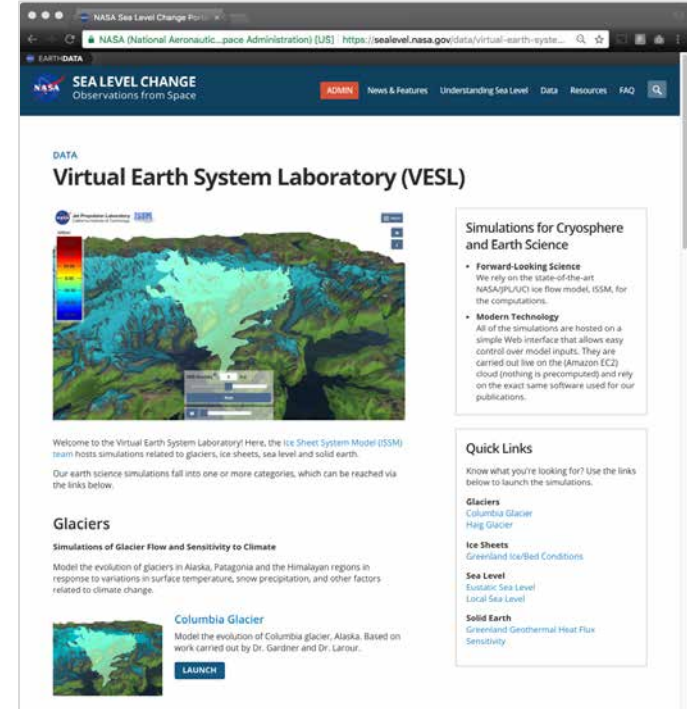
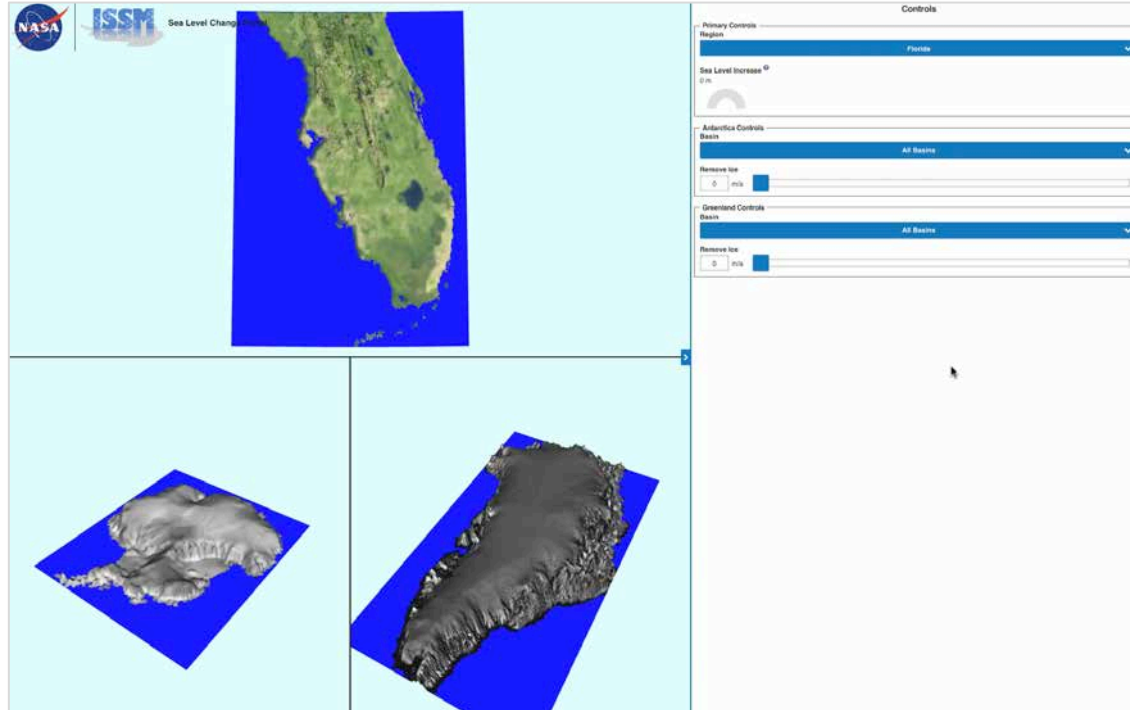


Local Sea Level



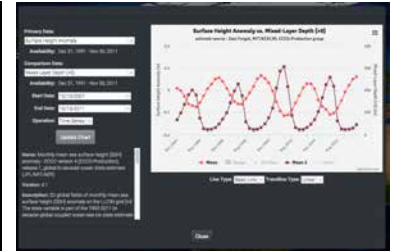
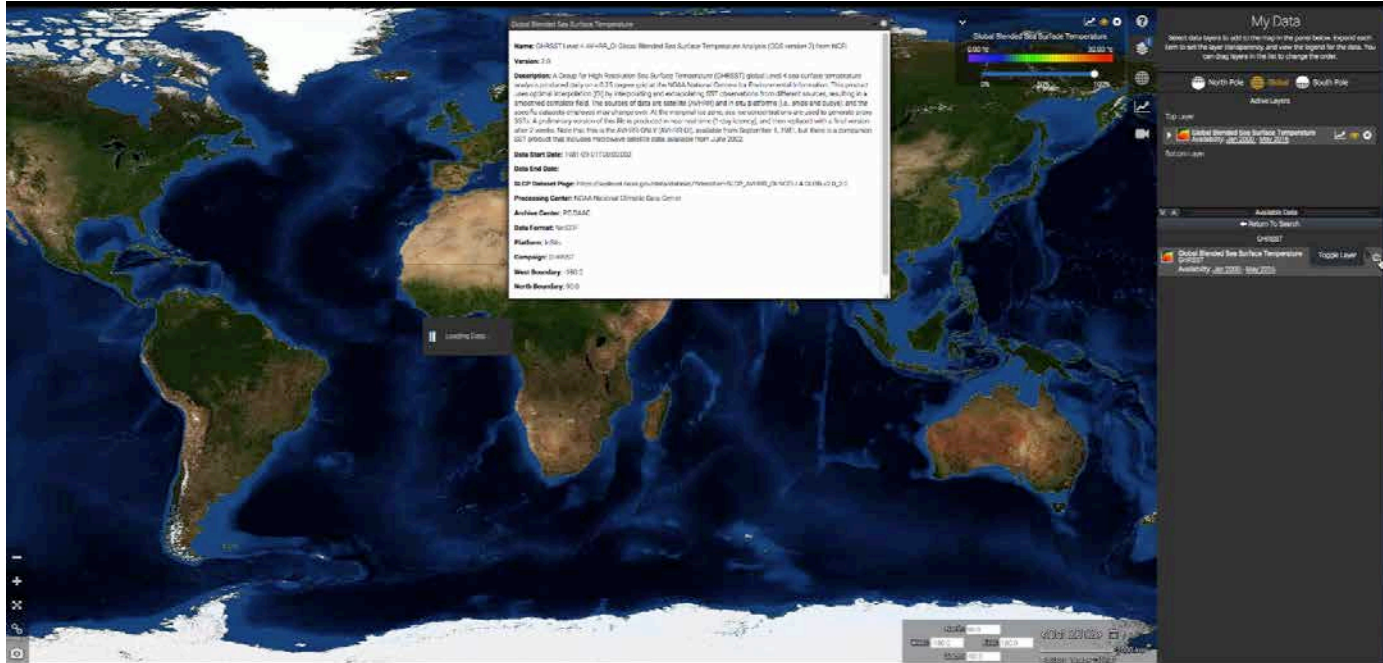
Greenland Steady-State-Friction

# Virtual Earth System Laboratory (VESL)



- Web-based 3D Simulations
- Computation on Amazon Cloud

# Analyze Sea Level On-The-Fly <https://sealevel.nasa.gov>

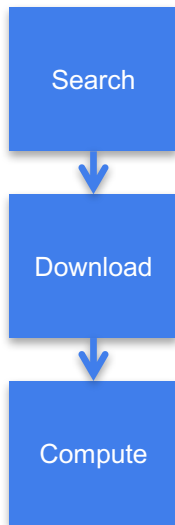


## Sea Level Change - Data Analysis Tool

Visualizations | Hydrological Basins | Time Series | Deseason | Data Comparison | Scatter Plot |  
 Latitude/Time Hovmöller | Etc.



# Traditional Method for Analyze Satellite Measurements



Depending on the data volume (size and number of files)

It could take many hours of download – (e.g. 10yr of observational data could yield thousands of files)

It could take many hours of computation

It could require expensive local computing resources

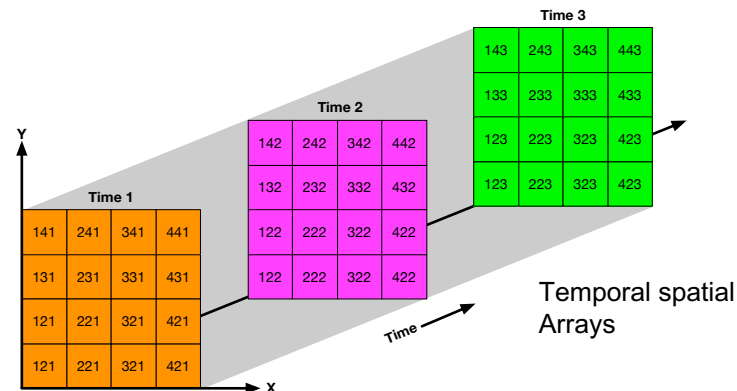
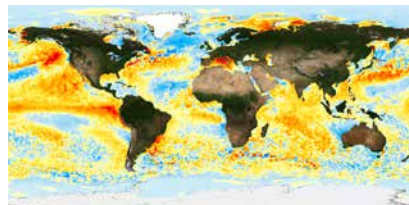
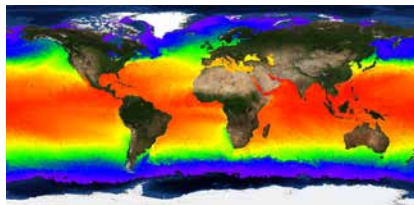
Often after result is produced, purge downloaded files

## Observations

Traditional methods for data analysis (time-series, distribution, climatology generation) can't scale to handle large volume, high-resolution data. They perform poorly

Performance suffers when involve large files and/or large collection of files

A high-performance data analysis solution must be free from file I/O bottleneck

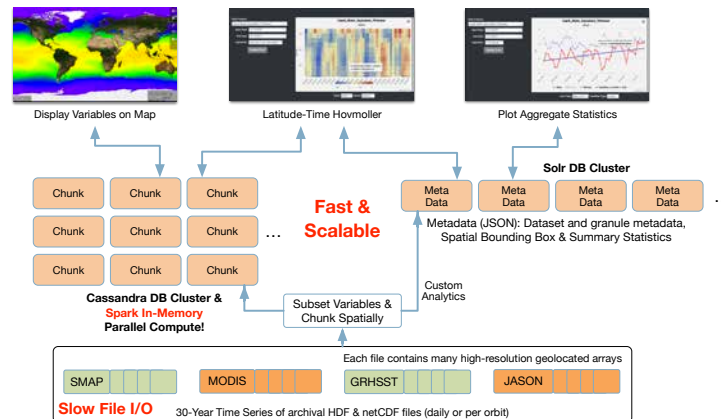


# NEXUS: Scalable Data Analytic Solution

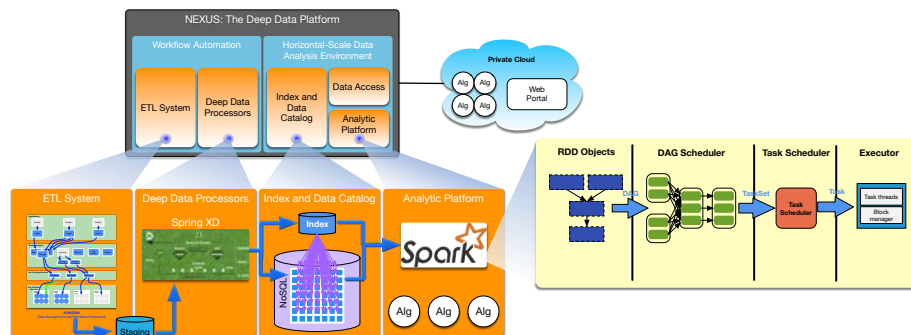
- NEXUS is a data-intensive analysis solution using a new approach for handling science data to enable large-scale data analysis
- Streaming architecture for horizontal scale data ingestion
- Scales horizontally to handle massive amount of data in parallel
- Provides high-performance geospatial and indexed search solution
- Provides tiled data storage architecture to eliminate file I/O overhead
- A growing collection of science analysis webservices using Apache Spark: parallel compute, in-memory map-reduce framework
- Pre-Chunk and Summarize Key Variables
  - Easy statistics instantly (milliseconds)
  - Harder statistics on-demand using Spark (in seconds)
  - Visualize original data (layers) on a map quickly (Cassandra store)
- Algorithms** – Time Series | Latitude/Time Hovmöller | Longitude/Time Hovmöller | Latitude/Longitude Time Average | Area Averaged Time Series | Time Averaged Map | Climatological Map | Correlation Map | Daily Difference Average

**Open Source: Apache License 2**

<https://github.com/dataplumber/nexus>



## Two-Database Architecture





# NEXUS Performance: Custom Spark vs. AWS EMR

**Dataset:** MODIS AQUA Daily

**Name:** Aerosol Optical Depth 550 nm (Dark Target) (MYD08\_D3v6)

**File Count:** 5106

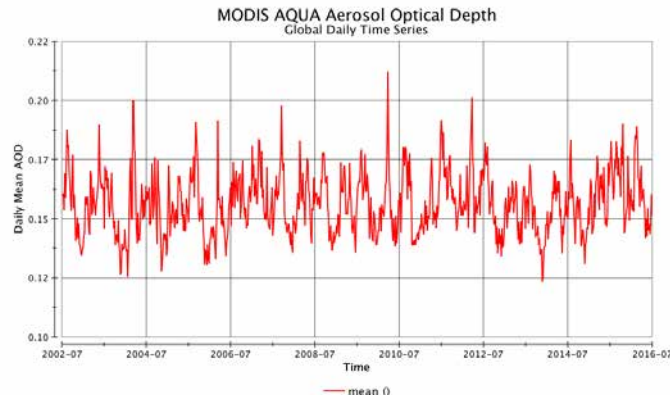
**Volume:** 2.6GB

**Time Coverage:** July 4, 2002 – July 3, 2016

**Giovanni:** A web-based application for visualize, analyze, and access vast amounts of Earth science remote sensing data without having to download the data.

- Represents current state of data analysis technology, by processing one file at a time
- Backed by the popular NCO library. Highly optimized C/C++ library

**AWS EMR:** Amazon's provisioned MapReduce cluster



Area Averaged Time Series on AWS - Boulder

July 4, 2002 - July 3, 2016

NEXUS Performance

Custom Spark vs. AWS EMR  
Ref. Speed - Giovanni: 1140.22 sec

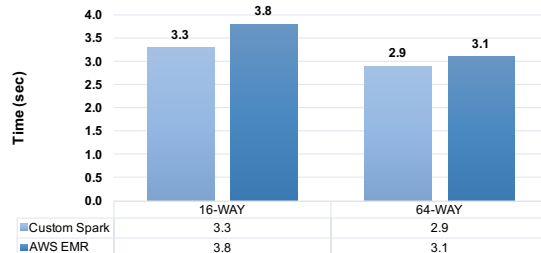


Area Averaged Time Series on AWS - Colorado

July 4, 2002 - July 3, 2016

NEXUS Performance

Custom Spark vs. AWS EMR  
Ref. Speed - Giovanni: 1150.6 sec

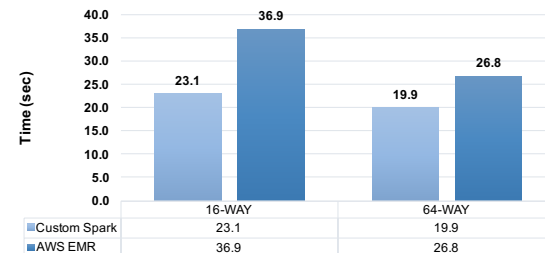


Area Averaged Time Series on AWS - Global

July 4, 2002 - July 3, 2016

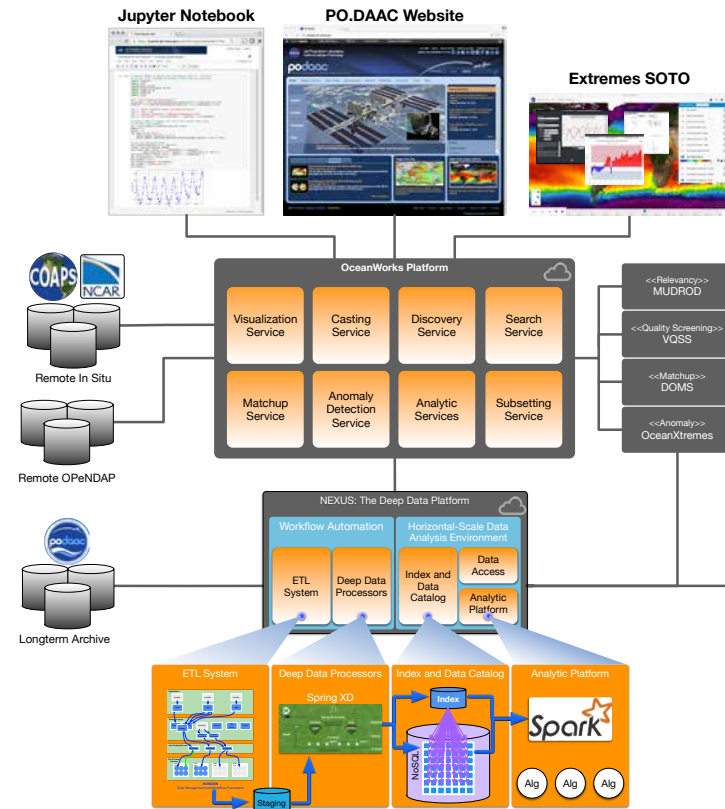
NEXUS Performance

Custom Spark vs. AWS EMR  
Ref. Speed - Giovanni: 1366.84 sec

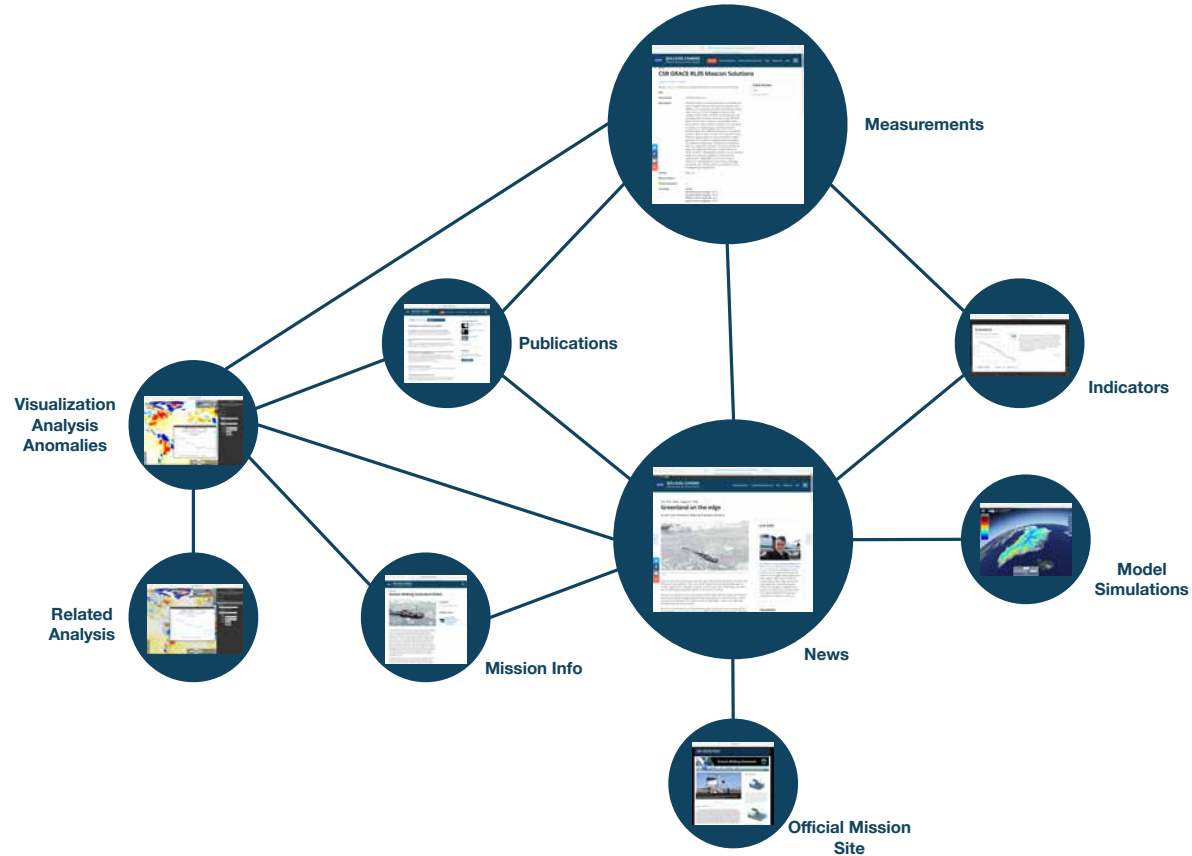


**OceanWorks** is to establish an **Integrated Data Analytic Center** at JPL by bringing together several big data technologies developed under the NASA ESTO/AIST program

- **OceanXtremes** – Anomaly detection and ocean science
- **NEXUS** – Big data analytic platform
- **Data Container Studies**
- **DOMS** – Distributed in-situ to satellite matchup
- **MUDROD** – Search relevancy and discovery – linking datasets, services, and anomalies through recommendations
- **VQSS** – Virtualized Quality Screening Service

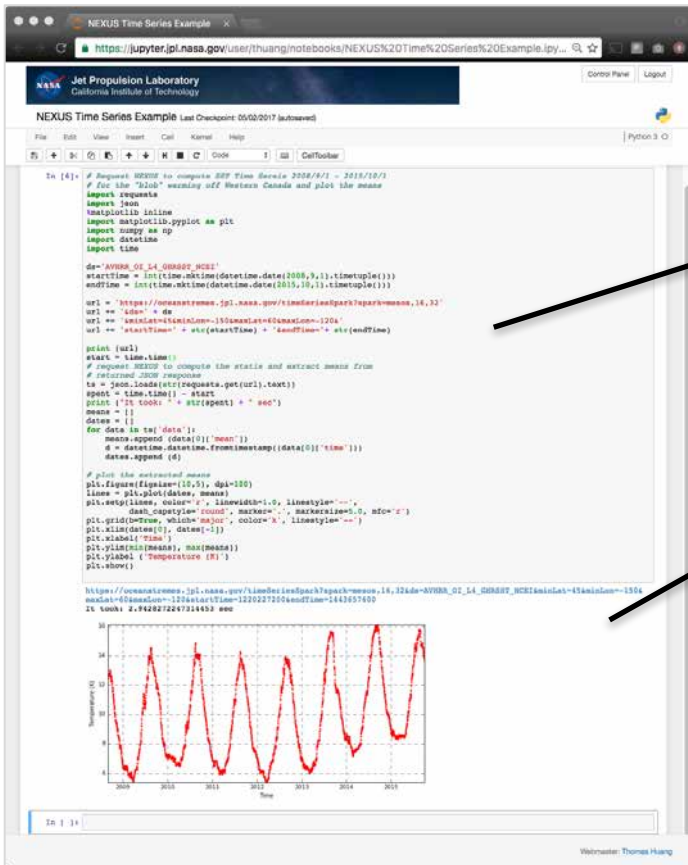


# Improve Information Discovery





# Enable Science without File Download



```
# Request NEXUS to compute SST Time Series 2008/9/1 - 2015/10/1
# for the "blob" warming off Western Canada and plot the means
...
ds='AVHRR_OI_L4_GHRSST_NCEI'

url = ... # construct the webservice URL request

# make request to NEXUS using URL request
# save JSON response in local variable
ts = json.loads(str(requests.get(url).text))

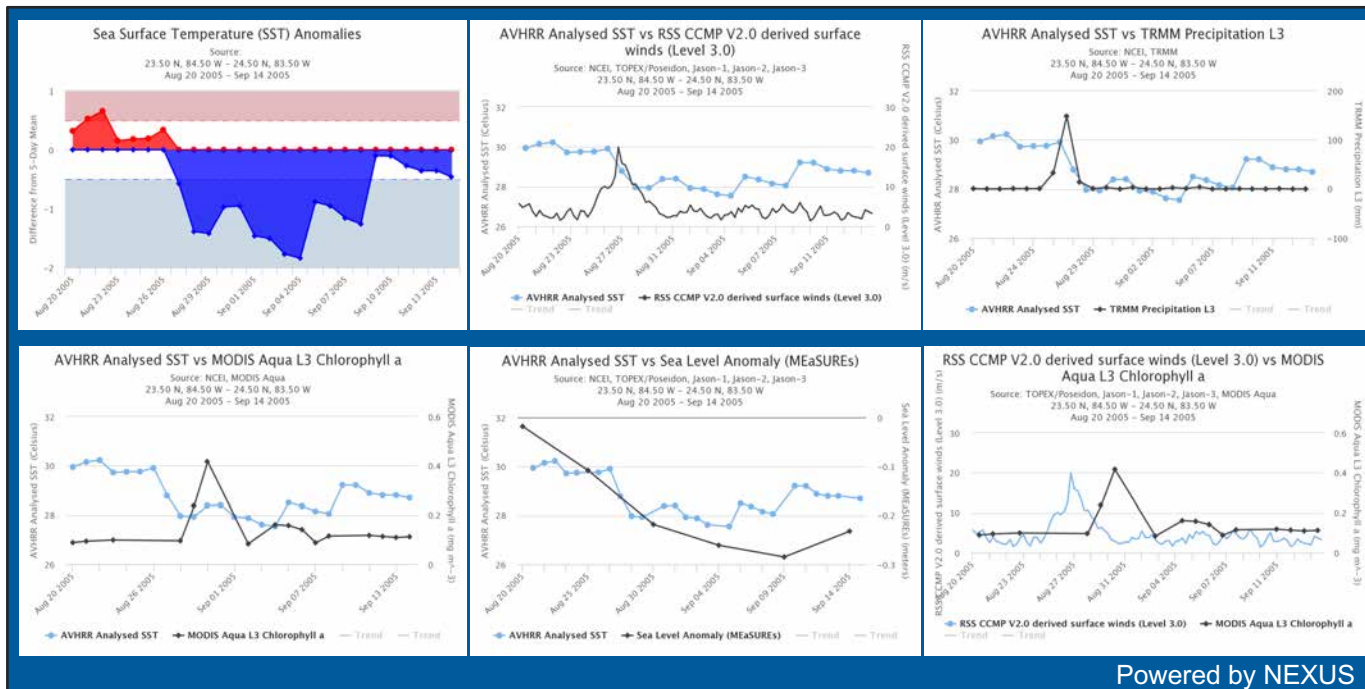
# extract dates and means from the response
means = []
dates = []
for data in ts['data']:
    means.append(data[0]['mean'])
    d = datetime.datetime.fromtimestamp((data[0]['time']))
    dates.append(d)

# plot the result
...
```

[https://oceanxtremes.jpl.nasa.gov/timeSeriesSpark?spark=mes,16,32&ds=AVHRR\\_OI\\_L4\\_GHRSST\\_NCEI&minLat=45&minLon=-150&maxLat=60&maxLon=-120&startTime=1220227200&endTime=1443657600](https://oceanxtremes.jpl.nasa.gov/timeSeriesSpark?spark=mes,16,32&ds=AVHRR_OI_L4_GHRSST_NCEI&minLat=45&minLon=-150&maxLat=60&maxLon=-120&startTime=1220227200&endTime=1443657600)

It took: 2.9428272247314453 sec

# Hurricane Katrina Study

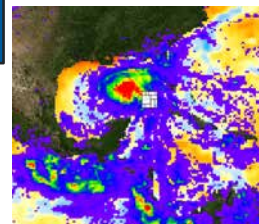


Powered by NEXUS

A study of a Hurricane Katrina-induced phytoplankton bloom using satellite observations and model simulations  
 Xiaoming Liu, Menghua Wang, and Wei Shi  
 JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, C03023, doi:10.1029/2008JC004934, 2009

Hurricane Katrina passed to the southwest of Florida on Aug 27, 2005. The ocean response in a 1 x 1 deg region is captured by a number of satellites. The initial ocean response was an immediate cooling of the surface waters by 2 °C that lingers for several days. Following this was a short intense ocean chlorophyll bloom a few days later. The ocean may have been “preconditioned” by a cool core eddy and low sea surface height.

The SST drop is correlated to both wind and precipitation data. The Chl-A data is lagged by about 3 days to the other observations like SST, wind and precipitation.



Hurricane Katrina  
 TRMM  
 overlay SST  
 Anomaly

Technology sharing through Free and Open Source Software (FOSS)

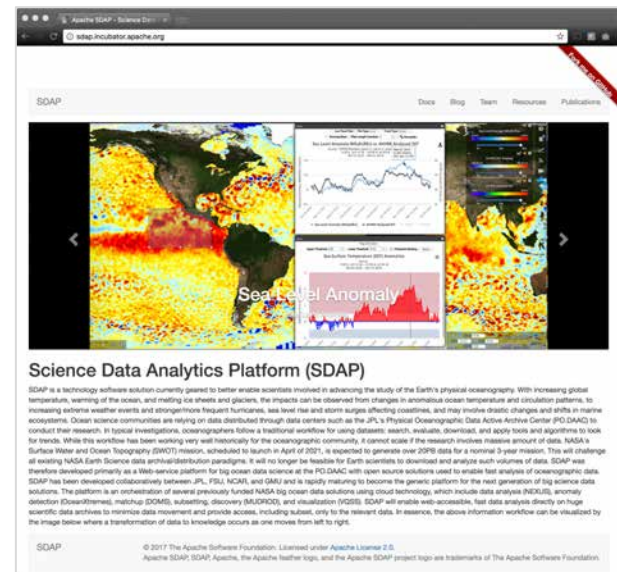
Further technology evolution that is restricted by projects / missions

**Science Data Analytic Platform (SDAP)**, the implementation of  
**OceanWorks**, in **Apache Incubator**

- Cloud platform
- Analyzing satellite and model data
- In situ data analysis and coordination with satellite measurements
- Fast data subsetting
- Mining of user interactions and data to enable discovery and recommendations
- Streamline deployment through container technology



<http://sdap.incubator.apache.org>





# In Summary

Traditional method for scientific research (search, download, local number crunching) is unable to keep up

Think beyond the archive

Connected information enables discovery

Community developed solution through open sourcing

Thanks to the NASA ESTO/AIST and Sea Level Rise programs, and the NASA ESDIS project

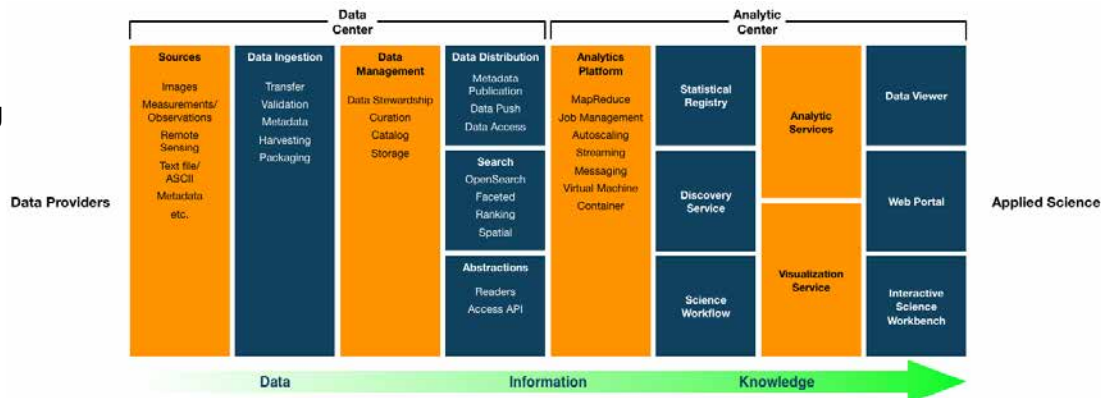
Investment in data and computational sciences

Data Centers need to be in the business of Enabling Science!

OceanWorks infusion 2018 – 2019

- Watch for changes to the Sea Level Change Portal
  - Even faster analysis capabilities
  - More variety of measurements – satellites, in situ, and models
  - Event more relevant recommendations
- NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC)

## Transforming Data to Knowledge





**National Aeronautics and  
Space Administration**

**Jet Propulsion Laboratory**  
California Institute of Technology  
Pasadena, California



**Thomas Huang**

Jet Propulsion Laboratory, California Institute of Technology